

Biological Modeling

Briefing paper for a Workshop
December 12, 1995
Wim Kimmerer

This briefing paper is intended to introduce the topic of the workshop and to provoke thought and discussion on some of the key issues. It is not a complete discussion of biological modeling, nor is it intended to steer the conclusions of the workshop. This paper has not been peer-reviewed, nor is it intended for publication.

What do we mean by modeling? Any concept or simplification of a real thing is a model. For the purposes of the workshop, however, we will confine the discussion to models with some (however limited) predictive capability. Such models can be arrayed along an axis of increasing empirical content in one direction, and increasing knowledge of mechanisms in the other. One end of the axis is anchored by purely empirical models (e.g. regression models). The other is held down by purely theoretical models. Models of greatest interest in the Bay-Delta scientific community are probably closer to the middle of this axis. They would be based on understanding of parts of the prototype system, have some connection to data from the prototype, and would be usable in predicting the response of the prototype to conditions not previously observed. Models that meet this description may not yet exist for the bay-delta ecosystem.

For the purposes of the workshop, we will also confine the discussion to models of predominantly open-water systems, as opposed to marshes and riparian zones. In addition, we will emphasize the use of models for management, while recognizing that models developed for research may ultimately have uses in management.

Why model? Models are constructed for a variety of purposes which can roughly be divided into research and management. Research models are constructed either to determine the consequences of a series of assumptions, or to translate data from one context to another. In many cases model predictions are compared with observations and the degree of "fit" of the model is evaluated. The most useful research models are those that do not do a good job of predicting, forcing researchers to reevaluate and revise assumptions. Unfortunately, in many cases an acceptable fit results in acceptance of the model by its author, with the result that alternative, equally plausible models are not tested.

Modeling for management has a different aim, which is to predict the consequences of management actions given the best estimate of how the system is believed to respond. Relatively few alternative methods are available to predict the effects of management activity or engineering changes to a natural system. Most often managers rely on expert opinion for guidance. In doing so, they fail to note that in fact they are relying on the experts' conceptual (or other) models, but with the assumptions not made explicit.

Constructing simulation models for management purposes has two advantages over the more traditional use of expert judgment: 1) assumptions are made explicit; and 2) the consequences of uncertainty in functions, parameters, and data can be explored. However, it is crucial to scale the complexity of the model to the level of understanding of the system being modeled. For example, it would make no sense to construct a detailed population model to predict the response of delta smelt to their environment until better knowledge of the controls on their abundance became available (although a research model might be useful in exploring those controls).

Model development can be quite expensive, especially where it must be combined with data collection to provide input to the model. However, for managing the bay-delta estuary and tributaries, modeling is essential for:

- Exploring the relative costs and benefits of different actions
- Investigating the consequences of alternative descriptions of the system
- Determining the key weaknesses in understanding of the system

Finally, modeling of biological systems is required by several statutes and regulations. Although these requirements may be somewhat naive and overly optimistic, they nevertheless require the bay-delta scientific community to examine carefully the opportunities for use of models in management.

A mismatch between expectations and capabilities One of the reasons to hold this workshop is a persistent mismatch between what managers and engineers expect of modeling, and what biologists think they can deliver. The high expectations of the engineering community arise from extensive experience with simulation of the physical environment (e.g. hydrology, hydrodynamics, temperature), which can give generally realistic predictions of physical conditions. Although models of hydrodynamics may seem quite complex, in fact they embody only one equation plus conservation of mass; their complexity arises mainly from the necessity to parameterize turbulent mixing at scales smaller than the length scale of the model cells.

Models of biological systems, on the other hand, must describe systems for which the equations are at best poorly known. To understand this, consider a model of a population of annually-reproducing fish. This model would describe the reproductive rate of the adult fish, and then the survival of the young fish as they grow to maturity. The population grows or decays at an annual rate equal to the product of the reproductive rate and the survival proportions for each life stage. At some point in the life cycle, there must be negative feedback or "density dependence", by which the survival is inversely related to population size. Without this feedback, the population will grow or decay without limit. Although there is plenty of evidence that such feedback exists, only rarely is it possible to determine the mechanism or even at what stage of the life cycle this feedback occurs. A model attempting to predict how the population will change as a result of changes in, say, egg survival will be completely unsuccessful if density-dependent mortality occurs in the larval stage.

As another complication, consider that most models of populations are trophic-dynamic, i.e. they describe changes in populations based mainly on changes in food supply (and sometimes predators). However, it is not clear that food supply is the principal influence on specific populations. In the Bay-Delta estuary, many populations vary positively with freshwater outflow, but this variation is probably not a result of covariation in food supply. A trophic-dynamic model would be unsuccessful at describing how these populations vary with their environment.

If biological models cannot be built from first principles, how can we proceed? It seems to me that a logical next step is to use the available data to construct models that incorporate known mechanisms along with some empirical information. This approach has been used for models of salmon smolt survival through the delta and striped bass production, although in both cases the models have been criticized on statistical grounds.

Examples of Models Modeling efforts to date in the bay-delta system have been constrained by the questions being asked (whether for management or research purposes) and the local emphasis on effects of freshwater flow. These constraints have led in the past to an emphasis on single-species models of striped bass and salmon, and less emphasis than elsewhere on lower trophic levels.

Some models currently in existence, listed by decreasing degree of empiricism, and increasing knowledge of mechanisms (names in parentheses are those presenting these models at the workshop)

- Fish-X2 models (Kimmerer)
- Salmon smolt survival (Williamson)
- Striped bass Young-of-the-year
- Survival models for threatened species (Botsford)
- "Particle"-tracking models (Quinn, Cowan)
- Salmon population models (Williamson)
- Striped bass individual-based population model (Cowan)
- South bay phytoplankton (Lucas)

Some models not now used in bay-delta:

- Rule-based simulation models
- Coupled physical-biological models (e.g. nutrient-phytoplankton-zooplankton models; these are used extensively to describe open-ocean systems)
- Material flow models (e.g. network models, trophic-dynamic models)
- Multi-species fishery models

Caveats for modeling Modelers and their employers need to keep the following in mind in developing, testing, and applying models:

1. *The form of the model depends on the questions* The first step in any modeling exercise should always be to decide what questions the model will attempt to answer. This step constrains the form, scale, and content of the model. A model built without specific questions in mind is unlikely to perform well at answering questions determined post hoc.
2. *The map is not the territory.* Modelers, with intimate knowledge of the content of their models, do not usually confuse model output with that of the prototype. Model users, on the other hand, can readily be persuaded that the model is an exact replica of the prototype with all of its complexity. It is incumbent on modelers to disabuse model users of the notion that the model is more than just a tool to be used in conjunction with other tools.
3. *Model validation requires an illogical statistical model* There is a lot of pseudo-rigor in the practice of calibrating a model against one set of data and "validating" it against another. Often this is done with some specific criteria, based on the measurements at hand, as to how well the model should fit the data, and sometimes with a statistical test of the model's fit to the data. However, the model prediction is the null hypothesis against which the data are tested. Statistical tests are designed to distinguish between the data and the null hypothesis, and the more data available, the more precise that distinction can be. Therefore, it is in the modeler's interest to have as few data as possible, with the widest possible confidence intervals, to insure that the model fits well. Collecting more data will practically guarantee deviation of model from data.
4. *Model validation is a flawed concept anyway* If we construct a model and "validate" it by comparing it with a set of data, what have we done? In fact, all validation does for us is give us a sense that model predictions are generally in the right ballpark. It does not permit us to infer that the model is a correct description of the system, or that model predictions for other sets of inputs would also be correct. There may be (and usually is) an infinite set of possible alternative models, many of which could fit the data better or be more true to the response of the prototype. Since the data set available for validation is

usually small and the number of parameters and functions large, some reasonable fit could be expected with a wide variety of alternative descriptions.

The future of modeling What will be the most fruitful paths for development of models useful in management over the next 5-10 years? This is a key question for the workshop; here I offer a few ideas for consideration.

Given the difficulties in developing realistic models of biological systems alluded to above, it is clear that any modeling effort will need to be realistic and clear about the uncertainties. This includes not only uncertainties in parameters and data (e.g. population indices), but also uncertainties in structure and function. Thus, models for management purposes should explicitly consider alternative formulations and display prominently the differences among alternatives.

Modeling will have to be done in close conjunction with research and monitoring programs. This has not always been the case in the past. Models have been developed by one group of people, and data collected by the other, with insufficient communication between groups. This results in a lack of "ownership" for the models, with the consequence that they are developed but not used. An alternative institutional framework may be needed to insure that models are integrated fully with other methods of investigation.

Consideration should be given to "meta-models", or models incorporating sub-models with different scales and levels of detail. For example, suppose a goal of a modeling exercise was to assess the changes in all estuarine-dependent species resulting from a specific change in the flow regime. A model could be constructed to answer this question specifically, but it would have limited applicability and flexibility. An alternative approach would be to have individual models of different subsystems (e.g. populations, races, regions), having different levels of complexity, and that could be queried by an overarching model. As understanding of the subsystems developed, the individual models could be revised without requiring a revision of the meta-model.

Finally, this workshop is a first step in bringing together scientists and engineers involved in modeling biological systems in the Bay-Delta. The current regulatory framework, and the interest in solving environmental problems of the Bay-Delta, suggest that modeling efforts will need to be intensified in the future. This process would benefit by frequent communication among interested parties. To the extent that this workshop is successful, it may be useful to continue holding conferences or workshops on related topics from time to time.